

Ensemble-based estimates of eigenvector error for empirical covariance matrices

DANE TAYLOR

*Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics,
University of North Carolina, Chapel Hill, NC 27599, USA; Statistical and Applied
Mathematical Sciences Institute, Research Triangle Park, NC 27709, USA*
dane.r.taylor@gmail.com

JUAN G. RESTREPO

Department of Applied Mathematics, University of Colorado, Boulder, CO, 80309, USA
juanga@colorado.edu

AND

FRANÇOIS G. MEYER

*Department of Electrical, Computer and Energy Engineering, University of Colorado,
Boulder, CO, 80309, USA*
fmeyer@colorado.edu

[Received on 30 December 2016]

Covariance matrices are fundamental to the analysis and forecast of economic, physical and biological systems. Although the eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{u}_i\}$ of a covariance matrix are central to such endeavors, in practice one must inevitably approximate the covariance matrix based on data with finite sample size n to obtain empirical eigenvalues $\{\hat{\lambda}_i\}$ and eigenvectors $\{\hat{\mathbf{u}}_i\}$, and therefore understanding the error so introduced is of central importance. We analyze eigenvector error $\|\mathbf{u}_i - \hat{\mathbf{u}}_i\|^2$ while leveraging the assumption that the true covariance matrix is drawn from an ensemble with known spectral properties, such as the distribution of and the gaps between eigenvalues. This approach complements previous analyses of eigenvector error that require the full set of eigenvalues to be known. In particular, we study the distribution of the expected square error $r = \mathbb{E}[\|\mathbf{u}_i - \hat{\mathbf{u}}_i\|^2]$ across the matrix ensemble for eigenvalues near a particular value λ and, for example, find for sufficiently large matrix size p and sample size n that the probability density of the squared error decays as $1/nr^2$. We support this and further results with numerical experiments.

Keywords: covariance matrix, empirical eigenvector, Wigner surmise, Wishart distribution.

Math Subject Classification: 37N99, 62B10, 94A17

1. Introduction

The spectral properties of covariance matrices are a central topic in mathematics, probability and statistics (Anderson [2003], Golub & Loan [2012], Hastie et al. [2009], Mehta [1991]) and provide a cornerstone to applications in physics, biology, economics and social science (Bassett et al. [2011], Delvenne et al. [2010], Elton et al. [2009], Gatti et al. [2010], Mantegna & Stanley [2000], Volkov et al. [2009], Weigt et al. [2009]). The estimation of eigenvectors of a sample covariance matrix remains a fundamental tool for these and numerous other application domains. Sample covariance matrices can be computed

locally if the dataset lies along a manifold, or globally if the data is organized along a linear subspace (Hastie et al. [2009]). Often, the practitioner has access to a generative stochastic model for the covariance matrix that can be derived from first principles or domain knowledge, and he/she needs to estimate the accuracy of eigenvectors calculated from a sample covariance matrix.

In this work, we are interested in the “classical” (large sample) framework, where one has access to n measurements of a p -dimensional vector \mathbf{x} and $n > p$. If we further assume that the sample covariance matrix $\tilde{\mathbf{C}}$ is distributed according to a Wishart distribution $W(\mathbf{C}, n)$ centered at population covariance matrix \mathbf{C} , then for a fixed p and $n \rightarrow \infty$, the expected error between a sample eigenvector $\tilde{\mathbf{u}}_i$ and the corresponding population eigenvector \mathbf{u}_i is asymptotically given for $i \in \{1, \dots, p\}$ by (Anderson [2003])

$$\mathbb{E} [n \|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2] \rightarrow \sum_{j=1; j \neq i}^p \frac{\lambda_j \lambda_i}{(\lambda_i - \lambda_j)^2}. \quad (1.1)$$

Here, $\tilde{\lambda}_1, \dots, \tilde{\lambda}_p$ and $\lambda_1, \dots, \lambda_p$ are the corresponding sample and population eigenvalues, respectively, which we assume to be simple and in ascending order. The main contributions of this paper are asymptotic $p \rightarrow \infty$ estimates, \hat{h}_i , of the quantities

$$h_i \triangleq \sum_{j=1; j \neq i}^p \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}, \quad (1.2)$$

when the eigenvalues $\lambda_1, \dots, \lambda_p$ are distributed according to a known limiting $p \rightarrow \infty$ spectral density $\rho(\lambda)$. These estimates can be used to bound the residual error between the sample and the population eigenvectors, $\mathbf{u}_i - \tilde{\mathbf{u}}_i$ in expectation, asymptotically for large n , using

$$\mathbb{E} [\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2] \approx \frac{1}{n} h_i. \quad (1.3)$$

The idea of taking advantage of existing *a priori* knowledge about the spectral density $\rho(\lambda)$ has led to novel insights and improved inference for covariance analyses (Bickel & Levina [2008], Lam & Fan [2009], Ledoit & P      [2011]). In practice, the probability distribution $\rho(\lambda)$ can be estimated from empirical data, or can sometimes be derived analytically (Kuhn [2008], Mehta [1991]). A situation of particular interest is when the covariance matrix follows a graphical (i.e., network-based) model in which complex network properties can give rise to different spectral densities (c.f., Benaych-Georges & Nadakuditi [2011], Chung et al. [2003], I. J. Farkas & Vicsek [2001], K.-I. Goh & Kim [2001], Peixoto [2013], S. N. Dorogovtsev & Samukhin [2003], Taylor et al. [2016a;b], Zhang et al. [2014]).

In order to estimate the size of the h_i in (1.2), we further use an approximation to the joint probability distribution $J(s^-, s^+)$ of the left and right gaps around each eigenvalue λ_i ,

$$s_i^+ \triangleq \lambda_{i+1} - \lambda_i \text{ and } s_i^- \triangleq \lambda_i - \lambda_{i-1}. \quad (1.4)$$

Finally, we note that because we are interested in computing estimates of h_i in (1.2) when both $p \rightarrow \infty$ and $n \rightarrow \infty$, we expect that the gaps between any two eigenvalues, s_i^+ and s_i^- , will go to zero. In fact, it is easy to show that the expected gap between two eigenvalues λ_i and λ_{i+1} should have size

$$\frac{1}{p\rho(\lambda_i)} \quad (1.5)$$

as $p \rightarrow \infty$. As a result, many of our results will be expressed in terms of the limiting gap given by (1.5).

Our main results provide estimates of the probability density function $f_H(h)$ of h_i in (1.2) in terms of λ_i , $\rho(\lambda_i)$, and p . Such bounds are of great consequence, because they provide the expected uncertainty associated with the sample eigenvectors, given a known limiting spectral density $\rho(\lambda)$, thereby providing an estimate of uncertainty across the entire ensemble of covariance matrices associated with $\rho(\lambda)$.

The paper is organized as follows. We state our main results in section 2. In section 3, we provide numerical simulations to validate our theoretical analysis. In section 4, we describe conditions in which (1.1), and thus our main results, are valid. The Appendix contains the derivations of our main results.

2. Main results

In this section, we present asymptotic ($n \rightarrow \infty$ and $p \rightarrow \infty$) approximations for the expected residual error of sample eigenvectors as well as their distribution across an ensemble of population covariance matrices. We first provide preliminary discussion in section 2.1. In sections 2.2–2.4, we present our three main results.

2.1 Definitions, notations and assumptions

In this work, we consider the $p \times p$ sample covariance matrix $\tilde{\mathbf{C}}$ computed from n observations, $\mathbf{x}_1, \dots, \mathbf{x}_n$, of a centered random vector $\mathbf{x} \in \mathbb{R}^p$. We denote by $\tilde{\lambda}_1 \leq \dots \leq \tilde{\lambda}_p$ the p sample eigenvalues of $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_p$ the corresponding sample eigenvectors. Likewise, $\lambda_1 \leq \dots \leq \lambda_p$ and $\mathbf{u}_1, \dots, \mathbf{u}_p$ denote the corresponding population eigenvalues and associated eigenvectors, respectively, of \mathbf{C} .

In order to provide estimates for h_i , we make the following technical assumptions.

Assumption 2.1. *We assume that the population eigenvalues are identically distributed with the limiting $p \rightarrow \infty$ spectral density $\rho(\lambda)$, and we assume weak convergence. We further assume that $\rho(\lambda)$ has compact support $[\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}^+$, and is continuous and differentiable on the interior of its support, $(\lambda_{\min}, \lambda_{\max})$.*

Many ensembles of symmetric random matrices satisfy Assumption 2.1 (see Anderson [2003], Kuhn [2008], Mehta [1991]) including, for example, those described by the semi-circle law (Pastur et al. [2011]). For some applications, it may also be beneficial to posit a parametric model for $\rho(\lambda)$, which can be estimated for small p and n and extended to the entire dataset.

Assumption 2.2. *We assume that the joint probability distribution $J(s^-, s^+)$ of the left and right gaps around each eigenvalue λ can be accurately approximated by*

$$J(s^-, s^+) \approx \frac{3^7 [p\rho(\lambda)]^5}{32\pi^3} [s^+ s^- (s^+ + s^-)] \exp\left(-\frac{[3p\rho(\lambda)]^2}{4\pi} [(s^+)^2 + (s^-)^2 + s^+ s^-]\right). \quad (2.1)$$

The joint distribution given by (2.1) was derived in Herman et al. [2007] using 3×3 matrices from the Gaussian Orthogonal Ensembles (GOE) and can be construed as a generalization of the Wigner surmise, which approximates marginal distributions for $J(s^-, s^+)$ and can be derived for 2×2 GOE matrices. As shown in Herman et al. [2007], (2.1) gives a very good approximation to the exact distribution. In addition, as demonstrated in our numerical simulations, (2.1) provides a good approximation for the ensemble of covariance matrices that we study (see Figure 2).

Importantly, since $J(s^-, s^+) = 0$ if $s^+ = 0$ or $s^- = 0$, assumption 2.2 implies that the p population eigenvalues are distinct, $\lambda_1 < \dots < \lambda_p$, and is akin to the “level repulsion of eigenvalues” observed in large random matrices (Bourgade et al. [2014]) that states that the gap probability is 0 for $s^\pm = 0$. Moreover, this assumption establishes $\mathcal{O}(1/p\rho(\lambda))$ as the asymptotic scaling for the expectation of s^\pm .

We may now state the main results of this work.

2.2 Main result 1: Estimate of h_i for large p

In the limit of large p and large n , we find the following approximation for h_i ,

$$\hat{h}_i = \lambda_i^2 \left[\left(\frac{1}{(s_i^-)^2} + \frac{1}{(s_i^+)^2} \right) + p\rho(\lambda_i) \left(\frac{1}{s_i^-} + \frac{1}{s_i^+} \right) \right]. \quad (2.2)$$

See Appendix A for the derivation.

We can explain the role of the different terms in (2.2) as follows. The left term in the squared brackets approximates the terms in (1.2) that involve the nearest neighbor eigenvalues of λ_i , which are respectively located at $\lambda_{i-1} = \lambda_i - s_i^-$ and $\lambda_{i+1} = \lambda_i + s_i^+$ and dominate the summation in (1.2) when p is large. This term does not require the knowledge of the probability distribution $\rho(\lambda)$. The second term accounts for the remaining terms in (1.2), which involve the remaining eigenvalues, $\{\lambda_j : |j - i| > 1\}$. Finally, as explained above, we recall that $1/p\rho(\lambda_i)$ is about the same order as the gap between two eigenvalues (Pastur et al. [2011]), and therefore all terms in the right-hand side of (2.2) can potentially obtain similar magnitudes.

We note that (2.2) allows one to approximate the expected error using (1.3)

$$\mathbb{E} [\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2] \approx \hat{h}_i/n \quad (2.3)$$

for the sample eigenvector $\tilde{\mathbf{u}}_i$ associated with eigenvalue λ_i of a large covariance matrix (i.e., $p \gg 1$) using knowledge of the population eigenvalue distribution $\rho(\lambda)$ and eigenvalue gaps s_i^\pm . Thus it does not require knowledge of the full set of sample eigenvalues $\{\tilde{\lambda}_i\}$. However, we stress that this approximation is valid only when p and n are sufficiently large, which we will explore in section 4.

2.3 Main result 2: Estimate of the probability density of h for large p

We again consider the case where p is large, and we study the limit of the distribution of \hat{h}_i for $p \rightarrow \infty$. By combining (2.2), and the expression of the joint probability for the eigenvalue gap given by (2.2), we obtain the following semi-analytical expression for the limiting probability density of the approximation \hat{h} to h ,

$$f_H(h) = - \int_{s^0(h)}^{\infty} J(s^*(h, s^+), s^+) \frac{\partial s^*(h, s^+)}{\partial h} ds^+, \quad (2.4)$$

where the variables $s^0(h)$ and $s^*(h, s^+)$ depend on the eigenvalue λ around which h is computed, and are respectively given by

$$s^0(h) = \frac{\lambda^2 p \rho(\lambda)}{2h} \left\{ 1 + \sqrt{1 + \frac{4h}{[\lambda p \rho(\lambda)]^2}} \right\}, \quad (2.5)$$

and

$$s^*(h, s^+) = \lambda^2 p \rho(\lambda) \frac{1 + \sqrt{1 + \frac{4}{[\lambda p \rho(\lambda)]^2} \left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p \rho(\lambda)}{s^+} \right)}}{2 \left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p \rho(\lambda)}{s^+} \right)}. \quad (2.6)$$

See Appendix B for the derivation.

The significance of (2.4) stems from the fact that it allows one to approximate the distribution of h and therefore the distribution of expected eigenvector errors, using the approximation (2.3), which again assumes sufficiently large p and n . Specifically, $f_H(h)$ estimates the distribution of expected residual error across the covariance-matrix ensemble associated—that is, as opposed to (1.1), which is an estimate for a single covariance matrix from the ensemble.

2.4 Main result 3: Asymptotic behavior of $f_H(h)$ for large h

Keeping λ and $\rho(\lambda)$ fixed, in the limit when the left gap, s^- , or right gap, s^+ , goes to zero, then h goes to infinity, and we find the following scaling behavior for the probability density function

$$f_H(h) = \mathcal{O}\left(\frac{p^2}{h^2}\right). \quad (2.7)$$

See Appendix C for the derivation. Therefore, $f_H(h)$ scales like h^{-2} for large h .

3. Numerical simulation results

We now report the results of numerical experiments to validate the theoretical predictions given by the main results described in section 2. We first describe the ensemble of covariance matrices used for these experiments.

3.1 Setting: Laplacian of k -regular graphs

We consider the following graphical model for covariance matrices. All the covariance matrices \mathbf{C} take the form

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T, \quad (3.1)$$

where \mathbf{X} is a random incidence matrix that describes the connectivity of a random graph $G = (V, E)$, with an arbitrary orientation of the edges. The entry $x_{e,v}$ of \mathbf{X} is given by

$$x_{e,v} = \begin{cases} 1 & \text{if } v \text{ is the head of the oriented edge } e, \\ -1 & \text{if } v \text{ is the tail of the oriented edge } e, \\ 0 & \text{otherwise.} \end{cases} \quad (3.2)$$

We generate the random incidence matrix for a k -regular graph using the configuration model (Newman [2003]). We note that the covariance matrix in (3.1) is equal to the (unnormalized) combinatorial Laplacian of a graph (Bapat [2010]).

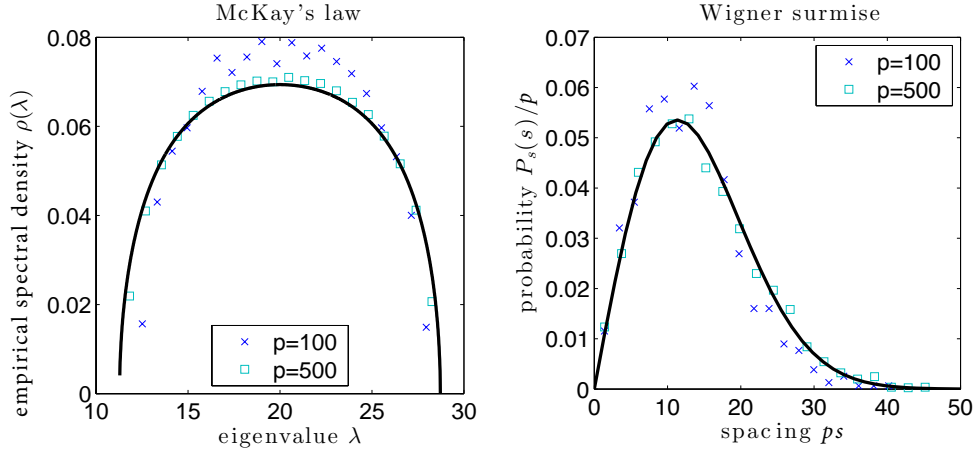


FIG. 1. Left: Empirical spectral density $\rho(\lambda)$ converges toward McKay's law (McKay [1981]) (black curve) in the limit of $p \rightarrow \infty$. Right: Distribution of normalized eigenvalue gaps $\{ps_i^+\}$ for eigenvalues $|\lambda_i - 20| < 1$ is well-described by the Wigner surmise (black curve) (González & Téllez [2008]).

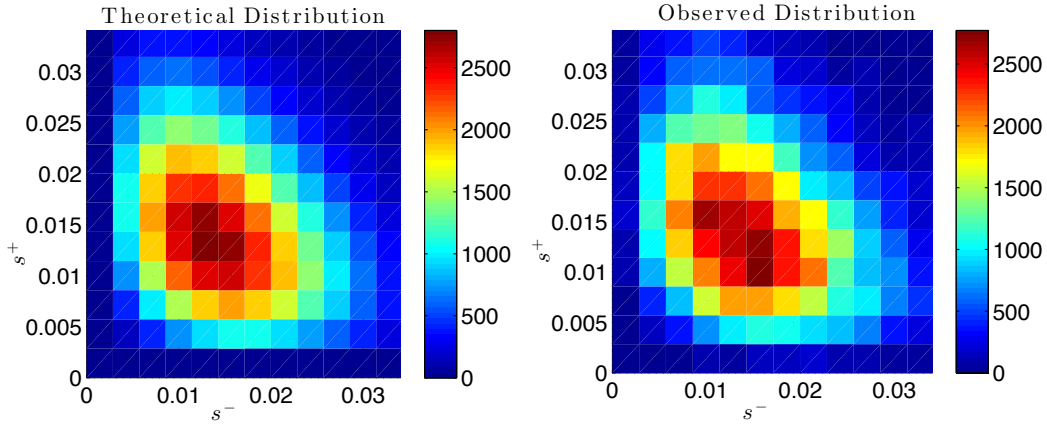


FIG. 2. Joint distribution of consecutive eigenvalue gaps $J(s^-, s^+)$: theoretical distribution (left) given by (2.1), and numerically observed distribution (right). The color indicates the unnormalized counting measure.

Figure 1-left displays the empirical spectral density computed over 50 realizations of covariance matrix \mathbf{C} , using matrix sizes $p = 100$ (blue crosses) and $p = 500$ (cyan squares). Note that p indicates both the matrix size and the number of vertices in the graph. As p increases from 100 to 500, we observe the convergence of the empirical spectral density toward the limiting density given by McKay's law (McKay [1981]) (black curve)

$$\rho(\lambda) = \begin{cases} \frac{k\sqrt{4(k-1) - \lambda^2}}{2\pi(k^2 - \lambda^2)} & \text{if } |\lambda| \leq 2\sqrt{k-1}, \\ 0 & \text{otherwise.} \end{cases}$$

McKay's law was originally obtained as the limiting spectral distribution for fixed k ; however, one

can also allow k to increase with p (Dumitriu & Pal [2012]). We also point out that in the numerical experiments to follow, we will estimate the distribution of eigenvalues directly from the data, rather than use McKay's law, as this approach would be more relevant for empirical data.

Figure 1-right displays the empirical probability density of the normalized spacing ps^+ for the set of eigenvalues $\{\lambda_i\}$ such that $|\lambda_i - 20| < 1$. We used approximately $2p\rho(\lambda)$ spacing values to compute the empirical densities. As expected, the eigenvalue gap distributions appear to be consistent with the Wigner surmise (González & Téllez [2008])

$$P(s) \approx \frac{\pi p^2 \rho^2(\lambda)}{2} s \exp\left(-\frac{\pi p^2 \rho^2(\lambda)}{4} s^2\right), \quad (3.3)$$

denoted by the black curve. Note that the agreement improves with increasing p .

To gain further insight into the gap distribution, and to validate Assumption 2.2, we compared the unnormalized counting measure with the joint eigenvalue-gap distribution $J(s^-, s^+)$ for the eigenvalues $\{\lambda_i\}$ such that $|\lambda_i - 20| < 1$. Figure 2-left displays the level sets of $p^2 J(s^-, s^+)$ according to (2.1) with $p = 1,000$. Figure 2-right shows the unnormalized counting measure computed across 100 covariance matrices of size $p = 1,000$. These are in very good agreement.

3.2 Experimental validation of main result 1

We first compared the estimate \hat{h}_i , given by (2.2), with the true values of h_i , defined by (1.2), for covariance-matrices ensemble described in section 3.1. We considered graphs of fixed degree $k = 20$ and $p = 100$ vertices.

In Figure 3-left, we compare (2.2) with the true value of h_i computed directly from the eigenvalues. The points lie close to the diagonal (dashed line), which validates the accuracy of the approximations. To illustrate the effect of the terms $p\rho(\lambda_i)\lambda_i^2/s_i^\pm$ in (2.2), we plot our approximation with (red plus symbols) and without (blue crosses) these corrections. One can observe that these terms improve the estimate for small h_i and have little effect for large h_i . This is expected since large h_i corresponds to very small s_i^\pm . In this limit, the correction terms become negligible as $(s_i^\pm)^{-2} \gg (s_i^\pm)^{-1}$.

In the next experiment, we compare \hat{h}_i given by (2.2) with a bootstrap estimate of the mean sample error, $n\hat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, for Wishart distribution $W(\mathbf{C}, n)$. Specifically, we generated a population covariance matrix \mathbf{C} with $k = 20$ and $p = 200$. We then generated 100 random realizations $\tilde{\mathbf{C}}$ from $W(\mathbf{C}, n)$ with $n = 10^7$. Let $\{\mathbf{u}_i\}_{i=1}^p$ be the eigenvectors of \mathbf{C} . For each random realization $\tilde{\mathbf{C}}$, we calculated its eigenvectors $\{\tilde{\mathbf{u}}_i\}_{i=1}^p$ and computed the residual error $\mathbf{u}_i - \tilde{\mathbf{u}}_i$ between the sample eigenvectors and the population eigenvectors. We then computed a bootstrap estimate, $\hat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, indicating the observed mean eigenvector error across the 100 realizations of $\tilde{\mathbf{C}}$.

In Figure 3-right, we plot the observed values $n\hat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ versus our prediction given by (2.3). The mean is plotted in black, and the standard deviation is shown in blue. We note that the solid curves lie very close to the diagonal indicating the accuracy of (2.3).

In these experiments, the sample size n was chosen to be sufficiently large so that (1.3) and (2.3) are accurate. We discuss in section 4 a simple and practical bound that can be used to choose appropriate values of n .

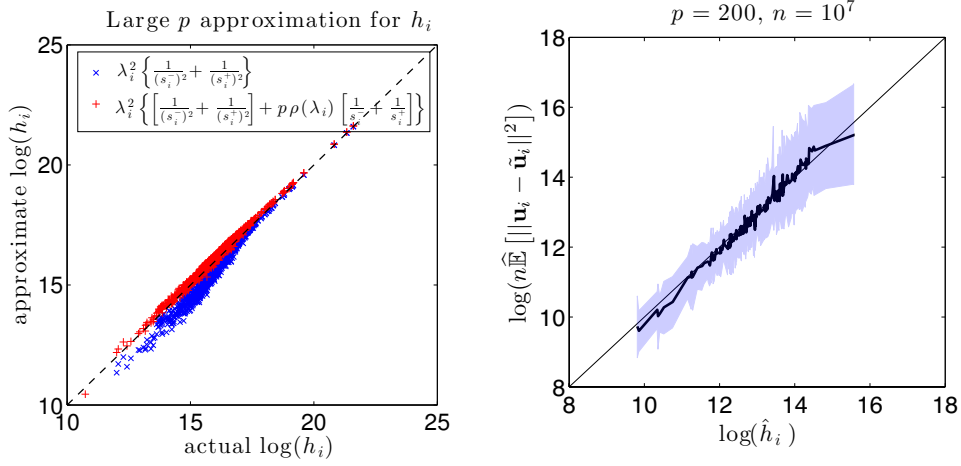


FIG. 3. Left: Approximation \hat{h}_i of $h_i = \sum_{j \neq i} \lambda_i \lambda_j / (\lambda_i - \lambda_j)^2$ given by (2.2) as function of the true h_i . We show \hat{h}_i with (red plus symbols) and without (blue crosses) the correction terms. Right: Bootstrap estimate of the sample mean error, $n\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, which is computed from 100 samples from Wishart distribution $W(\mathbf{C}, n)$ with $n = 10^7$, versus approximation \hat{h}_i given by (2.2). The mean is plotted by the black curve, and the standard deviation is shown in blue. See text for details.

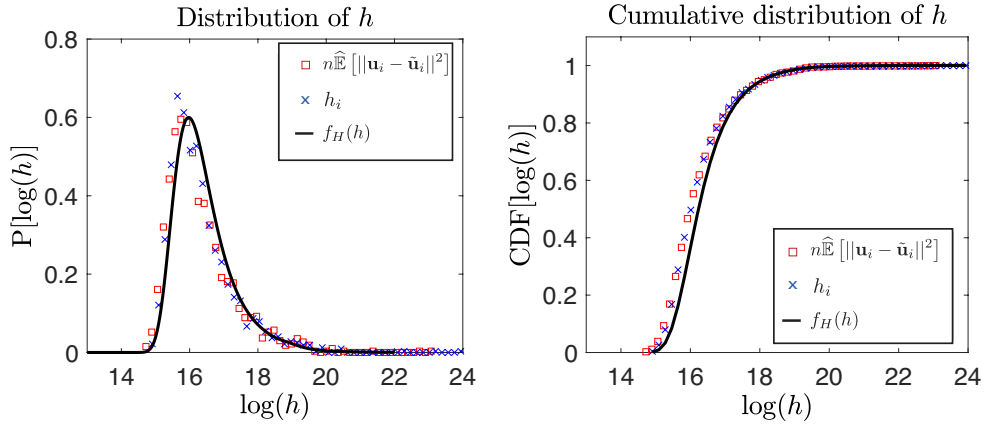


FIG. 4. Accuracy of probability density function (left) and cumulative density (right) of h for $n = 10^{10}$. We depict the following: (black curves) semi-analytical expression of the probability distribution $f_H(h)$ given by (2.4) and corresponding cumulative distribution; (blue crosses, \times): empirically observed distribution of h_i ; (red squares, \square): empirically observed distribution of bootstrap estimate, $\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$.

3.3 Experimental validation of main result 2

We now describe experiments that validate the second main result presented in section 2.3. We confirm that the approximation of $f_H(h)$ given by (2.4) is in good agreement with the empirical distribution of

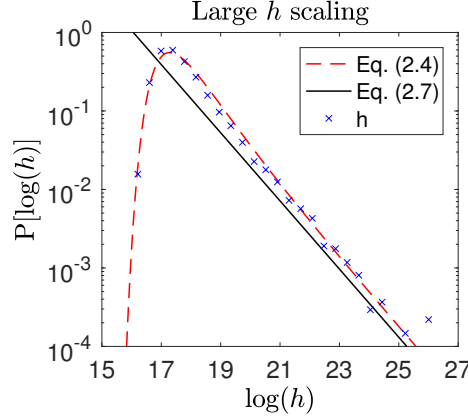


FIG. 5. Large- h scaling for $f_H(h)$. We plot $P[\log(h)]$ as a function of $\log(h)$: semi-analytical expression $f_H(h)$ computed from (2.4) in red (- -); limiting approximation to $f_H(h)$ for large h , given by (2.7) in black (-); empirical distribution, shown as blue crosses (\times).

h_i . Furthermore, we show experimentally that $f_H(h)$ in (2.4) also approximates the distribution of the expected residual error $\mathbb{E}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, provided that n and p are sufficiently large.

We generated 50 unweighted graphs of fixed degree $k = 20$ and fixed size $p = 1,000$. For each graph, we constructed the population covariance matrix, \mathbf{C} , as explained in section 3.1. For each \mathbf{C} , we generated 10 sample covariance matrices $\tilde{\mathbf{C}}$ from Wishart distribution $W(\mathbf{C}, n)$ with $n = 10^{10}$. For each $\tilde{\mathbf{C}}$, we calculated its eigenvectors $\{\tilde{\mathbf{u}}_i\}_{i=1}^p$ and computed the residual error, $\mathbf{u}_i - \tilde{\mathbf{u}}_i$, between the sample and population eigenvectors. We then computed a bootstrap estimate, $\hat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, of the mean sample error for each \mathbf{C} using the 10 realizations of $\tilde{\mathbf{C}}$.

In Figure 4-left, we use a solid black curve for the semi-analytical expression of the probability distribution $f_H(h)$ given by (2.4). We plot its corresponding cumulative distribution in Figure 4-right. We plot with blue crosses in both panels a numerically observed distribution of h_i , which we estimate using 50 covariances \mathbf{C} drawn from the graphical model described in section 3.1. We plot with red squares an empirical distribution of bootstrap estimates, $n\hat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$. As expected, the probability density function $f_H(h)$ provides a good approximation of the empirical distribution of h_i as well as the distribution of $n\hat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ (that is, provided n and p are both sufficiently large). However, we note that the distribution $f_H(h)$ is shifted slightly to the right.

3.4 Experimental validation of main result 3

We conclude with numerical validation of main result 3, $f_H(h) \propto h^{-2}$ for large h , which we presented in section 2.4.

We generated 500 covariance matrices \mathbf{C} using the graphical model described in section 3.1, with $k = 20$ and $p = 2,000$. Figure 5 displays $P[\log(h)]$ using our theoretical distribution $f_H(h)$ given by (2.4) (dashed red curve) as a function of $\log(h)$. We also display as a solid black line the limiting scaling behavior, $f_H(h) \propto h^{-2}$, given by (2.7). Finally, we compare these two probability density functions with the empirical distribution of $\log(h)$, shown as blue crosses. We note for large h that all distributions are parallel in this log-log plot, indicating that they have the same asymptotic power-law scaling.

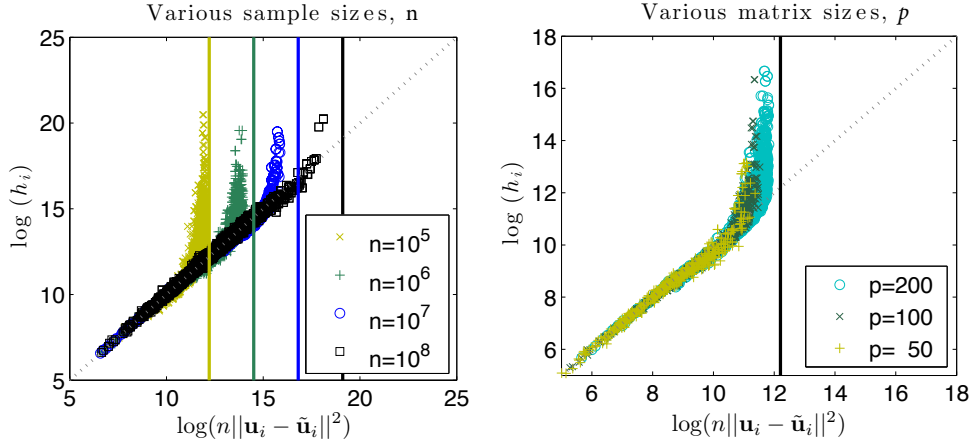


FIG. 6. True value of h_i given by (1.2) versus residual eigenvector error $n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2$ across 100 Wishart-distributed sample covariance matrices $\tilde{\mathbf{C}}$ with expectation \mathbf{C} with (left) $p = 200$ and various n , and (right) $n = 10^5$ and various p .

4. Discussion

A central motivator for our research has been equation (1.1), which describes the limiting $n \rightarrow \infty$ expected sample error $\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2$ of a sample eigenvector $\tilde{\mathbf{u}}_i$ for a covariance matrix drawn from a Wishart distribution. However, this equality only holds asymptotically. In this Discussion, we describe the conditions in which the approximation (1.3) is expected to be accurate. That is, when is the sample size n sufficiently large for given covariance matrix size p ?

The standard approach to this problem usually involves a tail bound. Instead, we use here a simple argument that yields a lower bound that works very well in practice. Indeed, we provide a necessary (but not sufficient) lower bound on n such that (1.1) and (2.2) are valid. Since both \mathbf{u}_i and $\tilde{\mathbf{u}}_i$ are normalized and we assume $\mathbf{u}_i \approx \tilde{\mathbf{u}}_i$, we have

$$\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2 = 2[1 - \langle \mathbf{u}_i, \tilde{\mathbf{u}}_i \rangle] \leq 2. \quad (4.1)$$

Under the approximation $\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2 \approx h_i/n$ given by (1.3), it follows that

$$n \geq h_i/2. \quad (4.2)$$

We now provide numerical support for this bound using the graphical model described in section 3.1. We first generate a k -regular graph with p vertices and compute the unnormalized Laplacian matrix, \mathbf{C} , which we treat as a covariance matrix. Let $\{\mathbf{u}_i\}_{i=1}^p$ be the eigenvectors of \mathbf{C} . In order to study the convergence of the empirical eigenvectors, we generate 100 random matrices $\tilde{\mathbf{C}}$ from the Wishart distribution $W(\mathbf{C}, n)$. For each random realization $\tilde{\mathbf{C}}$, we calculate its eigenvectors $\{\tilde{\mathbf{u}}_i\}_{i=1}^p$ and compute the residual error $\mathbf{u}_i - \tilde{\mathbf{u}}_i$ between the sample eigenvectors and the population eigenvectors.

Figure 6-left displays $\log(h_i)$ as a function of $\log(n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2)$ for each random realization of a Wishart matrix $\tilde{\mathbf{C}}$ for $p = 200$, $k = 5$, and several choices of n . For each value of n , we plot the bound given by (4.2), $\log(2n)$, as a vertical solid line. Figure 6-right displays a scatterplot of $\log(h_i)$ as a function of $\log(n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2)$ for $k = 5$, $n = 10^5$, and several values of p . We also plot $\log(2n)$ as a

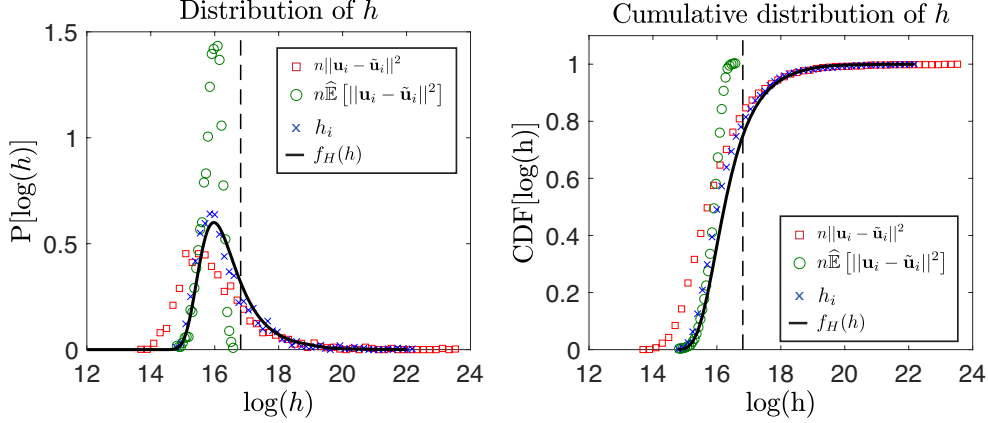


FIG. 7. Discrepancy between theoretical distribution $f_H(h)$ of h , given by (2.4), across a covariance matrix ensemble and a distribution of bootstrap estimates, $\widehat{\mathbb{E}}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, due to error for (1.1) and (1.3). We plot results for two sources of error. (green circles, \circ) Sample size n is too small and does not satisfy bound (4.2), which is shown by the vertical dashed lines. (red squares, \square): Insufficiently many (particularly, $R = 1$) samples are used to provide a reliable bootstrap estimate $\widehat{\mathbb{E}}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ to $\mathbb{E}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$. (See text for details.)

vertical solid line. Both panels illustrate (4.2) as a useful bound for considering when the approximation $h_i \approx \mathbb{E}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ given by (1.3) will be valid. Specifically, we require $\mathbb{E}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2] < 2$ and observe considerable discrepancy as $\mathbb{E}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2] \rightarrow 2$.

We conclude by exploring the effect of inaccuracy for (1.3) on the distribution of the bootstrap estimates, $n\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, of the mean residual error across the covariance matrix ensemble. We identify two sources of discrepancy: (i) choosing n too small so that the bound (4.2) is violated, and (ii) using insufficiently many samples from the Wishart distribution $W(\mathbf{C}, n)$ to provide an accurate bootstrap estimate, $n\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$. That is, the bootstrap estimate $\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ is only a reliable estimate of $\mathbb{E}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ if we generate enough random samples $\tilde{\mathbf{C}}$ from the Wishart distribution $W(\mathbf{C}, n)$.

We highlight these two sources of discrepancy with a numerical experiment similar to the one described in section 3.3. We generated 50 covariance matrices, \mathbf{C} , with $k = 20$ and $p = 1,000$, as explained in section 3.1. For each \mathbf{C} , we generated R random realizations $\tilde{\mathbf{C}}$ from the Wishart distribution $W(\mathbf{C}, n)$. For each $\tilde{\mathbf{C}}$, we calculated its eigenvectors $\{\tilde{\mathbf{u}}_i\}_{i=1}^p$ and computed the residual error $\mathbf{u}_i - \tilde{\mathbf{u}}_i$ between the sample eigenvectors and the population eigenvectors. We then computed a bootstrap estimate, $\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, of the mean sample error for each \mathbf{C} using the R realizations of $\tilde{\mathbf{C}}$.

In Figure 7-left and Figure 7-right, we plot as solid black curves the probability distribution $f_H(h)$ given by (2.4), and its corresponding cumulative distribution, respectively. We plot by blue crosses the empirical distribution of h_i , which we estimate using the 50 covariances \mathbf{C} . Note that $f_H(h)$ accurately predicts the observed distribution of h_i , since p is sufficiently large. In addition, we plot the distribution of the bootstrap estimates $\widehat{\mathbb{E}}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ of the mean sample error for $R = 10$ and $n = 10^7$ (green circles) as well as $R = 1$ and $n = 10^{10}$ (red squares). Note that when $R = 1$, the bootstrap estimate $\widehat{\mathbb{E}}[\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ is actually just the sample error $\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2$.

Observe that both distributions disagree with $f_H(h)$ for different reasons: For $R = 10$ and $n = 10^7$,

the distribution of $\widehat{\mathbb{E}}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ is expected to differ because $n = 10^7$ is too small and does not satisfy the bound given by (4.2) (vertical dashed lines). On the other hand, for $R = 1$ and $n = 10^{10}$, the sample error does not provide a good bootstrap estimate for the mean sample error $\mathbb{E}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$, which is the relevant quantity that is described in (1.1) and (1.3). We observe in Figure 7 that using too few samples (i.e., small R) affects the distribution of $\widehat{\mathbb{E}}[n\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2]$ by shifting it toward small values of h (see red squares).

Acknowledgment

The work of D. T. was partially supported by NSF Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute (SAMSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

- Anderson, T. W. (2003) *An Introduction to Multivariate Statistical Analysis*. John Wiley and Son, Inc., Hoboken, New Jersey, 3rd ed.
- Bapat, R. B. (2010) *Graphs and Matrices*. Springer.
- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M. & Grafton, S. T. (2011) Dynamic reconfiguration of human brain networks during learning. *Proceeding of the National Academy of Sciences*, **108**(18), 7641–7646.
- Benaych-Georges, F. & Nadakuditi, R. R. (2011) The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, **227**(1), 494–521.
- Bickel, P. J. & Levina, E. (2008) Covariance regularization by thresholding. *The Annals of Statistics*, pages 2577–2604.
- Bourgade, P., Erdős, L. & Yau, H.-T. (2014) Edge universality of beta ensembles. *Communications in Mathematical Physics*, **332**(1), 261–353.
- Chung, F., Lu, L. & Vu, V. (2003) Spectra of random graphs Spectra of random graphs with given expected degrees. *Proceeding of the National Academy of Sciences, USA*, **100**, 6313–6318.
- Delvenne, J.-C., Yaliraki, N., Sophia, N. & Barahona, M. (2010) Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, **107**(29), 12755–12760.
- Dumitriu, I. & Pal, S. (2012) Sparse regular random graphs: Spectral density and eigenvectors. *The Annals of Probability*, **40**(5), 2197–2235.
- Elton, E. J., Gruber, M. J., Brown, S. J. & Goetzmann, W. N. (2009) *Modern portfolio theory and investment analysis*. John Wiley and Sons.
- Gatti, D. M., Barry, W. T., Nobel, A. B., Rusyn, I. & Wright, F. A. (2010) Heading down the wrong pathway: on the influence of correlation within gene sets. *BMC Genomics*, **11**(1), 574.
- Golub, G. H. & Loan, C. F. V. (2012) *Matrix Computations*, volume 3. JHU Press.
- González, D. L. & Téllez, G. (2008) Wigner Surmise for Domain Systems. *Journal of Statistical Physics*, **132**(1), 187–205.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning*, volume 2. New York: Springer.
- Herman, D., Ong, T. T., Usaj, G., Mathur, H. & Baranger, H. U. (2007) Level Spacings in Random Matrix Theory and Coulomb Blockade Peaks in Quantum Dots. *Physical Review B*, **76**, 195448.
- I. J. Farkas, I. Derényi, A.-L. B. & Vicsek, T. (2001) Beyond the semicircle Beyond the semicircle law. *Physical Review E*, **64**, 026704.
- K.-I. Goh, B. K. & Kim, D. (2001) Spectra and eigenvectors of scale-free networks. *Physical Review E*, **64**, 051903.
- Kuhn, R. (2008) Spectra of sparse random matrices. *Journal of Physics A*, **41**, 295002.

- $$h_i^+ = \frac{\lambda_i \lambda_{i+1}}{(\lambda_i - \lambda_{i+1})^2} + \sum_{j=i+2}^p \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}. \quad (\text{A.5})$$

We study the large p behavior of (A.4) and (A.5) by separately considering the nearest-neighbor terms and the summations. In particular, we will obtain approximations that rely only on the right and left nearest-neighbor eigenvalue gaps,

$$s_i^\pm = |\lambda_i - \lambda_{i\pm 1}|. \quad (\text{A.6})$$

We first consider the isolated terms,

$$\frac{\lambda_i \lambda_{i\pm 1}}{(\lambda_i - \lambda_{i\pm 1})^2} = \frac{\lambda_i (\lambda_i \pm s_i^\pm)}{(s_i^\pm)^2} \quad (\text{A.7})$$

$$= \frac{\lambda_i^2}{(s_i^\pm)^2} [1 + \mathcal{O}(s_i^\pm)]. \quad (\text{A.8})$$

Using that $s_i^\pm \rightarrow 0$ as $p \rightarrow \infty$ (which is established by assumption 2.2 and convergences, in expectation, with rate $s_i^\pm = \mathcal{O}(1/p)$), we find the asymptotic estimate

$$\frac{\lambda_i \lambda_{i\pm 1}}{(\lambda_i - \lambda_{i\pm 1})^2} \rightarrow \frac{\lambda_i^2}{(s_i^\pm)^2}. \quad (\text{A.9})$$

We now turn our attention to the summations, which we will approximate using the limiting $p \rightarrow \infty$ spectral density $\rho(\lambda)$ of the normalized empirical counting measure of the eigenvalues. More precisely, consider a sequence of size- p symmetric covariance matrices, each having eigenvalues $\{\lambda_j\}$ for $j \in \{1, \dots, p\}$. We define for each matrix the empirical spectral density

$$\rho_p(\lambda) = p^{-1} \sum_j \delta(\lambda_j), \quad (\text{A.10})$$

where $\delta(\lambda)$ is the Dirac delta function and $\lambda \in \mathbb{R}$. We assume the covariance matrices are drawn from an ensemble such that the sequence $\{\rho_p(\lambda)\}$ weakly converges, implying that

$$\int_{-\infty}^{\infty} \rho_p(\lambda) f(\lambda) d\lambda \rightarrow \int_{-\infty}^{\infty} \rho(\lambda) f(\lambda) d\lambda \quad (\text{A.11})$$

as $p \rightarrow \infty$ for any continuous and bounded function $f(\lambda)$. We assume $\rho(\lambda)$ is continuous, bounded, has compact support (denoted $\text{supp}(\rho)$), and is differentiable on $\text{supp}(\rho)$. For notational convenience, we assume $\text{supp}(\rho) = (\alpha, \beta)$ for some $\alpha, \beta \in \mathbb{R}$, allowing us to replace the limits of integration in (A.11) by (α, β) . However, our analysis can be easily extended to unions of such intervals.

We begin by rewriting the summations in (A.4) and (A.5) as the integration of function

$$f_{\lambda_i}(\lambda) = \frac{\lambda_i \lambda}{(\lambda_i - \lambda)^2}. \quad (\text{A.12})$$

with probability measure $\rho_p(\lambda)$ given by (A.10),

$$\frac{1}{p} \sum_{j=1}^{i-2} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} = \int_{\alpha}^{\lambda_{i-1}} \rho_p(\lambda) f_{\lambda_i}(\lambda) d\lambda, \quad (\text{A.13})$$

$$\frac{1}{p} \sum_{j=i+2}^p \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} = \int_{\lambda_{i+1}}^{\beta} \rho_p(\lambda) f_{\lambda_i}(\lambda) d\lambda. \quad (\text{A.14})$$

Because $f_{\lambda_i}(\lambda)$ is unbounded at the singularity $\lambda = \lambda_i$, (A.11) does not describe the behavior of integral $\int_{\alpha}^{\beta} \rho_p(\lambda) f_{\lambda_i}(\lambda) d\lambda$, which we find to diverge with p for any $\lambda_i \in \text{supp}(\rho)$. Fortunately, (A.13) and (A.14) do not require integration across the singularity at $\lambda = \lambda_i$; however, the limits of integration, i.e., λ_{i-1} in (A.13) and λ_{i+1} in (A.14), depend on p (and converge to the singularity at λ_i). Thus, (A.11) is also not directly applicable to (A.13) and (A.14).

To proceed, we restrict our attention to (A.13) since analogous results can be obtained for (A.14). We consider, for the moment, (A.13) with fixed upper limit $\lambda_i - \varepsilon$ for $\varepsilon > 0$ and $\varepsilon \approx 0$. Equation (A.11) implies the $p \rightarrow \infty$ limit

$$\int_{\alpha}^{\lambda_i - \varepsilon} f_{\lambda_i}(\lambda) \rho_p(\lambda) d\lambda \rightarrow \int_{\alpha}^{\lambda_i - \varepsilon} f_{\lambda_i}(\lambda) \rho(\lambda) d\lambda. \quad (\text{A.15})$$

We now study how the right-hand side of (A.15) scales with ε . Using that both $\rho(\lambda)$ and $f_{\lambda_i}(\lambda)$ are differentiable for $\lambda \in \text{supp}(\rho) \setminus \{\lambda_i\}$, we implement integration by parts, treating the numerator and denominator separately, to obtain

$$\int_{\alpha}^{\lambda_i - \varepsilon} f_{\lambda_i}(\lambda) \rho(\lambda) d\lambda = \lambda_i \frac{(\lambda_i - \varepsilon) \rho(\lambda_i - \varepsilon)}{\varepsilon} - \lambda_i \int_{\alpha}^{\lambda_i - \varepsilon} \frac{\rho(\lambda) + \lambda \rho'(\lambda)}{\lambda_i - \lambda} d\lambda. \quad (\text{A.16})$$

The first term in the right hand side of (A.16) has the $\varepsilon \rightarrow 0$ asymptotic estimate

$$\lambda_i \frac{(\lambda_i - \varepsilon) \rho(\lambda_i - \varepsilon)}{\varepsilon} \rightarrow \frac{\lambda_i^2 \rho(\lambda_i)}{\varepsilon}. \quad (\text{A.17})$$

The second term on the right-hand side of (A.16) is bounded as

$$\begin{aligned} \left| \lambda_i \int_{\alpha}^{\lambda_i - \varepsilon} \frac{[\rho(\lambda) + \lambda \rho'(\lambda)]}{\lambda_i - \lambda} d\lambda \right| &\leq \lambda_i \left[\sup_{\lambda \in (\alpha, \lambda_i - \varepsilon]} |\rho(\lambda) + \lambda \rho'(\lambda)| \right] \int_{\alpha}^{\lambda_i - \varepsilon} \frac{1}{|\lambda_i - \lambda|} d\lambda \\ &= \lambda_i \left[\sup_{\lambda \in (\alpha, \lambda_i - \varepsilon]} |\rho(\lambda) + \lambda \rho'(\lambda)| \right] \ln \left(\frac{\lambda_i - \alpha}{\varepsilon} \right) \end{aligned} \quad (\text{A.18})$$

It follows that the second term in the right-hand side of (A.16) has scaling $\mathcal{O}(\ln(1/\varepsilon))$ and is dominated in the limit $\varepsilon \rightarrow 0$ by the first term, which is $\mathcal{O}(1/\varepsilon)$. We combine (A.17) and (A.18) to obtain the $\varepsilon \rightarrow 0$ asymptotic estimate

$$\int_{\alpha}^{\lambda_i - \varepsilon} f_{\lambda_i}(\lambda) \rho(\lambda) d\lambda \rightarrow \frac{\lambda_i^2 \rho(\lambda_i)}{\varepsilon}. \quad (\text{A.19})$$

We finally note that in the case where $\rho'(\lambda)$ is unbounded, it is straightforward to separate the integral on the left-hand side of (A.18) into two domains, one containing all values λ where $\rho'(\lambda)$ is unbounded and the second domain having upper limit $\lambda_i - \varepsilon$. The first integral will converge to zero due to (A.11), whereas the second satisfies the bound given by (A.18), implying that the integral term in (A.16) is $\mathcal{O}(\ln(1/\varepsilon))$ provided that $\rho(\lambda)$ is differentiable in a small neighborhood containing λ_i .

We study the $p \rightarrow \infty$ limiting behavior for the right-hand side of (A.13) by considering the following identity,

$$\int_{\alpha}^{\lambda_{i-1}} f_{\lambda_i}(\lambda) \rho_p(\lambda) d\lambda = \int_{\alpha}^{\lambda_i - s_i^-} f_{\lambda_i}(\lambda) \rho(\lambda) d\lambda + \int_{\alpha}^{\lambda_i - s_i^-} f_{\lambda_i}(\lambda) [\rho_p(\lambda) - \rho(\lambda)] d\lambda. \quad (\text{A.20})$$

The first term on the right-hand side grows linearly with p , which is straightforward to show by setting $\varepsilon = s_i^-$ in (A.19) and using that $s_i^- = \mathcal{O}(1/p)$. Turning our attention to the second term on the right-hand side of (A.20), recall that it would converge to zero if the upper limit of integration was fixed. However, $\lambda_i - s_i^-$ limits to λ_i and the $p \rightarrow \infty$ behavior of the second term therefore depends on the rate of weak convergence for $\rho_p(\lambda) \rightarrow \rho(\lambda)$. We assume that this term scales sublinearly with p and is therefore dominated by the first term on the right-hand side of (A.20). Under this assumption (and by conducting a similar analysis for (A.5)), we obtain the asymptotic estimates

$$p \int_{\alpha}^{\lambda_{i-1}} f_{\lambda_i}(\lambda) p \rho_p(\lambda) d\lambda \rightarrow \frac{\lambda_i^2 p \rho(\lambda_i)}{s_i^-}, \quad (\text{A.21})$$

$$p \int_{\lambda_{i+1}}^{\beta} f_{\lambda_i}(\lambda) \rho_p(\lambda) d\lambda \rightarrow \frac{\lambda_i^2 p \rho(\lambda_i)}{s_i^+}. \quad (\text{A.22})$$

In summary, we combine (A.21), (A.22) and (A.17) to obtain the asymptotic large p approximation,

$$h_i^{\pm} \approx \frac{\lambda_i^2}{(s_i^{\pm})^2} + \frac{p \rho(\lambda_i) \lambda_i^2}{s_i^{\pm}}, \quad (\text{A.23})$$

which gives the approximation (2.2). We stress that this approximation assumes a sufficiently high rate of weak convergence for the spectral density so that the second term on the right-hand side of (A.20) is sublinear.

B. Appendix B. Derivation of main result 2

In this section, we take a different perspective, and consider h_i , defined by (1.2), to be the realization of a random variable that is a function of the corresponding family of random covariance matrices. Using the approximation \hat{h} of h provided by (2.2), we derive an estimate for the probability distribution, $P(h)$ of h . Let us denote by H the random variable for which h_i is a realization.

Our goal is to remove the dependency on the random variables s^+ and s^- in (2.2), so that \hat{h} becomes a function of only λ , which is distributed according the density $\rho(\lambda)$. The only missing ingredients are the probability distributions of s^+ and s^- . We note that these two random variables are correlated, and thus our line of attack involves using an approximation to the joint probability for the eigenvalue gaps, $J(s^-, s^+)$, and derive an expression for the limiting probability density of the approximation \hat{h} . In this section, we keep the discussion general, and derive an expression that is valid for all $\rho(\lambda)$.

We assume that the joint probability distribution $J(s^-, s^+)$ of the left and right gaps around each eigenvalue λ can be approximated by (2.1), which is reproduced below for ease of presentation,

$$J(s^-, s^+) \approx \frac{3^7 [p \rho(\lambda)]^5}{32 \pi^3} [s^+ s^- (s^+ + s^-)] \exp \left(- \frac{[3 p \rho(\lambda)]^2}{4 \pi} [(s^+)^2 + (s^-)^2 + s^+ s^-] \right).$$

The expression (2.1) was derived in Herman et al. [2007] using 3×3 matrices from the Gaussian Orthogonal Ensembles (GOE). As suggested by our numerical simulations (see Figure 2), (2.1) provides a good approximation for the covariance matrices that we study.

To derive the distribution $P(h)$ of H , we first consider the cumulative distribution

$$F(h) \triangleq P(H < h). \quad (\text{A.1})$$

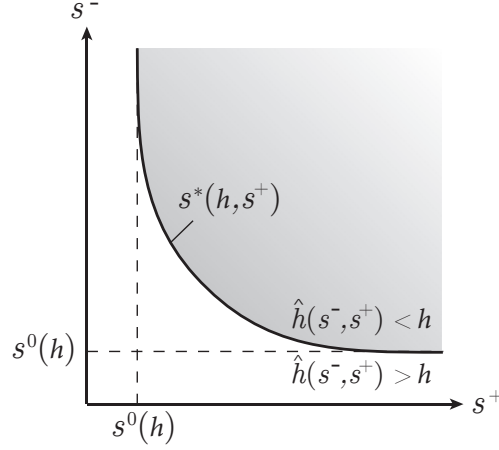


FIG. A.8. The cumulative distribution $F(h)$ for the random variable H is shown as the integral of $J(s^-, s^+)$ over region \mathcal{S} given by (A.2) (shaded region). This region corresponds to $s^+ \in (s^0(h), \infty)$ and $s^- \in (s^*(h, s^+), \infty)$, where $s^0(h)$ is found so that $h^+(s^+) < h$ for $s^+ > s^0(h)$ and $s^*(h, s^+)$ is found so that $\hat{h}(s^-, s^+) < h$ for $s^- > s^*(h, s^+)$.

Given an eigenvalue λ , we can find all the pairs of gaps s^- and s^+ , such that \hat{h} in (2.2) is less than h . Let

$$\mathcal{S} \triangleq \{(s^-, s^+) : \hat{h}(s^-, s^+) < h\} \quad (\text{A.2})$$

be this set. We then proceed to compute the measure of \mathcal{S} using the joint probability density function defined above,

$$F(h) = \int_{\mathcal{S}} J(s^-, s^+) ds^- ds^+. \quad (\text{A.3})$$

It turns out that we can describe analytically the set \mathcal{S} (see Figure A.8). For a given value of h , $\hat{h}(s^-, s^+) < h$ implies that both $h^+ < h$ and $h^- < h$, where h^\pm is given by (A.23) (with the subscript omitted), and therefore the region of integration has the lower bounds $s^- > s^0(h)$ and $s^+ > s^0(h)$, where $s^0(h)$ is given by

$$\begin{aligned} s^0(h) &= \frac{\lambda^2 p\rho(\lambda)}{2h} + \sqrt{\frac{\lambda^2}{h} + \left(\frac{\lambda^2 p\rho(\lambda)}{2h}\right)^2} \\ &= \frac{\lambda^2 p\rho(\lambda)}{2h} \left(1 + \sqrt{1 + \frac{4h}{[\lambda p\rho(\lambda)]^2}}\right), \end{aligned} \quad (\text{A.4})$$

which follows directly from solving (A.23) for s^\pm with $h^\pm = h$. We therefore integrate s^+ over the range $(s^0(h), \infty)$. For given values h and s^+ , requiring that $\hat{h}(s^-, s^+) > h$ implies that $s^- > s^*(s^+, h)$, where $s^*(h, s^+)$ is found by substituting $h \mapsto \hat{h}(s^-, s^+)$ in (2.2) and solving for the positive root of s^- ,

$$s^*(h, s^+) = \lambda^2 p\rho(\lambda) \frac{1 + \sqrt{1 + \frac{4}{[\lambda p\rho(\lambda)]^2} \left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p\rho(\lambda)}{s^+}\right)}}{2 \left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p\rho(\lambda)}{s^+}\right)}. \quad (\text{A.5})$$

We therefore integrate s^- over the range $(s^*(h, s^+), \infty)$,

$$F(h) = \int_{s^0(h)}^{\infty} \int_{s^*(h, s^+)}^{\infty} J(s^-, s^+) ds^- ds^+. \quad (\text{A.6})$$

To obtain an estimate for the distribution of h , $f_H(h)$, we differentiate (A.6) with respect to h to obtain

$$\begin{aligned} f_H(h) &= \frac{\partial}{\partial h} \int_{s^0(h)}^{\infty} \int_{s^*(h, s^+)}^{\infty} J(s^-, s^+) ds^- ds^+ \\ &= -\frac{\partial s^0}{\partial h}(h) \int_{s^*(h, s^0(h))}^{\infty} J(s^-, s^0(h)) ds^- + \int_{s^0(h)}^{\infty} \frac{\partial}{\partial h} \left[\int_{s^*(h, s^+)}^{\infty} J(s^-, s^+) ds^- \right] ds^+ \\ &= -\int_{s^0(h)}^{\infty} J(s^*(h, s^+), s^+) \frac{\partial s^*(h, s^+)}{\partial h} ds^+. \end{aligned} \quad (\text{A.7})$$

We note that in the above derivation, the first term in the second line vanishes since $s^*(h, s^+) \rightarrow \infty$ in the limit $s^+ \rightarrow s^0(h)$ and $J(s^-, s^+)$ is bounded.

C. Appendix C. Derivation of main result 3

With h distributed according to $f_H(h)$, given by (A.7), we derive in this section an asymptotic expression for $f_H(h)$ in the limit $h \rightarrow \infty$. Examining (2.2), we note that $\hat{h}(s^-, s^+)$ is large when s^- and/or s^+ are small, and thus in the large h limit one can consider only the contributions of the terms proportional to s_-^{-2} and s_+^{-2} ,

$$h \approx \frac{\lambda^2}{(s^-)^2} + \frac{\lambda^2}{(s^+)^2}. \quad (\text{A.1})$$

In this case, we find

$$s^0(h) = \frac{\lambda}{\sqrt{h}}, \quad (\text{A.2})$$

$$s^*(h, s^+) = \frac{\lambda s^+}{[(s^+)^2 h - \lambda^2]^{1/2}}, \quad (\text{A.3})$$

$$\frac{\partial}{\partial h} (s^*(h, s^+)) = \frac{-\lambda (s^+)^3}{2[(s^+)^2 h - \lambda^2]^{3/2}} = \frac{-1}{2\lambda^2} [s^*(h, s^+)]^3. \quad (\text{A.4})$$

Substituting these values into (A.7) and dropping the arguments for s^* , i.e. $s^*(h, s^+) \mapsto s^*$, we find

$$\begin{aligned} f_H(h) &= -\int_{\sqrt{\lambda^2/h}}^{\infty} \left(\frac{3^7 [p\rho(\lambda)]^5}{32\pi^3} s^+ s^* (s^* + s^+) e^{-\frac{[3p\rho(\lambda)]^2}{4\pi} [(s^*)^2 + (s^+)^2 + s^* s^+]} \right) \left(\frac{-(s^*)^3}{2\lambda^2} \right) ds^+ \\ &= \frac{3^7 [p\rho(\lambda)]^5}{32\pi^3} \frac{1}{2\lambda^2} \int_{\sqrt{\lambda^2/h}}^{\infty} \left((s^+ (s^*)^5 + (s^+)^2 (s^*)^4) e^{-\frac{[3p\rho(\lambda)]^2}{4\pi} [(s^*)^2 + (s^+)^2 + s^* s^+]} \right) ds^+. \end{aligned}$$

The change of variables $u = (s^+)^2 h - \lambda^2$ transforms this into

$$f_H(h) = \frac{3^7 [p\rho(\lambda)]^5}{32\pi^3} \frac{\lambda^2}{4} h^{-7/2} I(h), \quad (\text{A.5})$$

where we have defined

$$I(h) = \int_0^\infty \left(1 + \frac{\lambda^2}{u}\right)^{5/2} (u^{1/2} + \lambda) e^{-\varphi(u)/h} du, \quad (\text{A.6})$$

and

$$\varphi(u) = \frac{[3p\rho(\lambda)]^2}{4\pi} (u + \lambda^2) \left(1 + \frac{\lambda}{\sqrt{u}} + \frac{\lambda^2}{u}\right). \quad (\text{A.7})$$

The distribution $f_H(h)$ in (A.5) depends on h through the power law $h^{-7/2}$ as well as $I(h)$. In Appendix D, we show that (A.6) has the large- h scaling $I(h) = \mathcal{O}\left(\frac{h^{3/2}}{p^3}\right)$. Combining this with (A.5), we find $f_H(h) = \mathcal{O}\left(\frac{p^2}{h^2}\right)$ for large h .

D. Appendix D. Large- h scaling of $I(h)$

We now study how $I(h)$ given by (A.6) scales in the limit of large h . Recall that the limit of large h corresponds to when an eigenvalue λ_i has a nearest-neighboring eigenvalue that is very close (i.e., $|\lambda_i - \lambda_{i\pm j}| \ll 1$), which results in large values of h_i and subsequently the error of the empirical eigenvector (i.e., large $\|\mathbf{u}_i - \tilde{\mathbf{u}}_i\| \approx h_i/n$).

Our strategy for evaluating (A.6) is to split the integral into three regions of integration, which are chosen based on studying the function $\varphi(u)$. Examining (A.7) for limiting values of u , we find that the function $\varphi(u)$ approaches $+\infty$, both as $u \rightarrow 0$ and as $u \rightarrow \infty$, and has the minimum

$$\min_{u \in [0, \infty)} \varphi(u) = \varphi(\lambda^2) = \frac{27}{2\pi} [\lambda p\rho(\lambda)]^2, \quad (\text{A.1})$$

which occurs at $u = \lambda^2$. For large h , there are two values of u such that $\varphi(u) = h$. We refer to these values as $u_1(h)$ and $u_2(h)$, with $u_1(h) < u_2(h)$. Considering the limits $u \rightarrow 0$ and $u \rightarrow \infty$, we find the asymptotic approximations

$$u_1(h) \rightarrow \frac{[3p\rho(\lambda)]^2}{4\pi} \lambda^4 h^{-1}, \quad (\text{A.2})$$

$$u_2(h) \rightarrow \frac{4\pi}{[3p\rho(\lambda)]^2} h. \quad (\text{A.3})$$

We will evaluate (A.6) by dividing the integration into three ranges, $I(h) = I_1(h) + I_2(h) + I_3(h)$, where we define

$$I_1(h) = \int_0^{u_1(h)} \left(1 + \frac{\lambda^2}{u}\right)^{5/2} (u^{1/2} + \lambda) e^{-\varphi(u)/h} du, \quad (\text{A.4})$$

$$I_2(h) = \int_{u_1(h)}^{u_2(h)} \left(1 + \frac{\lambda^2}{u}\right)^{5/2} (u^{1/2} + \lambda) e^{-\varphi(u)/h} du, \quad (\text{A.5})$$

$$I_3(h) = \int_{u_2(h)}^\infty \left(1 + \frac{\lambda^2}{u}\right)^{5/2} (u^{1/2} + \lambda) e^{-\varphi(u)/h} du. \quad (\text{A.6})$$

We now study the $h \rightarrow \infty$ scaling for integrals $I_1(h)$, $I_2(h)$, and $I_3(h)$. Beginning with (A.4), we first note that for the range $u \in (0, u_1(h)]$ that

$$(u + \lambda^2)^{5/2}(u^{1/2} + \lambda) \leq (u_1 + \lambda^2)^{5/2}(u_1^{1/2} + \lambda). \quad (\text{A.7})$$

It follows that

$$\left(1 + \frac{\lambda^2}{u}\right)^{5/2} (u^{1/2} + \lambda) \leq E u^{-5/2}, \quad (\text{A.8})$$

where

$$E(\lambda) = (u_1(h) + \lambda^2)^{5/2}(u_1(h)^{1/2} + \lambda). \quad (\text{A.9})$$

Note that $E(\lambda) \approx \lambda^6$ as $h \rightarrow \infty$, since $u_1(h) \rightarrow 0$. Similarly, since u is positive, one finds

$$\begin{aligned} \varphi(u) &= \frac{[3p\rho(\lambda)]^2}{4\pi} (u + \lambda^2)[1 + (\lambda^2/u) + (\lambda^2/u)^{1/2}] \\ &\geq \frac{[3p\rho(\lambda)]^2}{4\pi} (\lambda^2)(\lambda^2/u) \\ &= F u^{-1}, \end{aligned} \quad (\text{A.10})$$

where we have defined

$$F = \frac{[3p\rho(\lambda)]^2}{4\pi} \lambda^4. \quad (\text{A.11})$$

Using these two inequalities, we have

$$I_1(h) \leq E(\lambda) \int_0^{u_1(h)} u^{-5/2} e^{-F/(hu)} du, \quad (\text{A.12})$$

$$= E(\lambda) \left(\frac{h}{F}\right)^{3/2} \int_{F/(hu_1(h))}^{\infty} w^{1/2} e^{-w} dw \quad (\text{A.13})$$

which uses the change of variables $w = F/(hu(h))$. Using (A.2), the lower limit of integration converges as $F/(hu_1(h)) \rightarrow 1$ with $h \rightarrow \infty$. The integral in (A.13) therefore limits to a constant, implying that $I_1(h)$ is dominated by a term which scales like $h^{3/2}$.

To estimate $I_3(h)$, note for large h that (A.3) implies $u > \lambda^2$ for any $u > u_2(h)$. It follows that

$$\left(1 + \frac{\lambda^2}{u}\right)^{5/2} (u^{1/2} + \lambda) \leq 2^{5/2}(2u^{1/2}). \quad (\text{A.14})$$

The integral $I_3(h)$ is thus dominated by

$$I_3(h) \leq 8 \int_{u_2(h)}^{\infty} u^{1/2} e^{-\varphi(u)/h} du, \quad (\text{A.15})$$

$$\leq 8 \int_{u_2(h)}^{\infty} u^{1/2} \exp\left(-\frac{[3p\rho(\lambda)]^2}{4\pi} \frac{u}{h}\right) du, \quad (\text{A.16})$$

where the second inequality uses $u > 0$ and $\lambda^2/u > 0$ to bound

$$\begin{aligned} \varphi(u) &= \frac{[3p\rho(\lambda)]^2}{4\pi} (u + \lambda^2)[1 + (\lambda^2/u) + (\lambda^2/u)^{1/2}] \\ &\geq \frac{[3p\rho(\lambda)]^2}{4\pi} u. \end{aligned} \quad (\text{A.17})$$

We define the change of variables $w = \frac{[3p\rho(\lambda)]^2}{4\pi} \frac{u}{h}$ to obtain

$$I_3(h) \leq 8 \left(\frac{[3p\rho(\lambda)]^2}{4\pi} \right)^{3/2} \int_{\frac{[3p\rho(\lambda)]^2}{4\pi} u_2(h)/h}^{\infty} w^{1/2} e^{-w} dw. \quad (\text{A.18})$$

From (A.3), the lower limit of integration converges as $\frac{[3p\rho(\lambda)]^2}{4\pi} u_2(h)/h \rightarrow 1$ and the integral in (A.18) converges to a constant as $h \rightarrow \infty$. Therefore $I_3(h)$ is also bounded by a term scaling as $h^{3/2}$.

We will now show that $I_2(h)$ has scaling $\mathcal{O}(h^{3/2})$ (as opposed to the other terms, which we showed are bounded by terms that scale as $h^{3/2}$). Note that because of our definition of u_1 and u_2 , and using that $\varphi(u)$ reaches a minimum at $u = \lambda^2 \in (u_1, u_2)$, we find the bounds

$$\varphi(\lambda^2)/h \leq \varphi(u)/h \leq 1 \quad (\text{A.19})$$

for any $u \in (u_1, u_2)$. Substituting these into (A.6), we bound $I_2(h)$ as

$$Q(h)e^{-1} \leq I_2(h) \leq Q(h)e^{-\varphi(\lambda^2)/h}, \quad (\text{A.20})$$

where we have defined

$$Q(h) \equiv \int_{u_1(h)}^{u_2(h)} \left(1 + \frac{\lambda^2}{u} \right)^{5/2} (u^{1/2} + \lambda) du. \quad (\text{A.21})$$

Using the asymptotic approximations for $u_1(h)$ and $u_2(h)$ given by (A.2) and (A.3), we integrate (A.21) using the software Mathematica (using the “Series[$Q(h)$, {h, Infinity, 1}]” command) to obtain its asymptotic behavior,

$$Q(h) \approx \frac{2^4 \pi^{3/2}}{3^4 [p\rho(\lambda)]^3} h^{3/2}. \quad (\text{A.22})$$

Furthermore, we combine $\varphi(\lambda^2)/h \rightarrow 0$ with (A.20) to obtain the asymptotic bound

$$Q(h)e^{-1} \leq I_2(h) \leq Q(h). \quad (\text{A.23})$$

We combine (A.23) with (A.13) and (A.18) to obtain the large- h scaling $I(h) = \mathcal{O}\left(\frac{h^{3/2}}{p^3}\right)$.